

# PATENT COOPERATION TREATY

From the  
INTERNATIONAL SEARCHING AUTHORITY

To:  
HADASSA WATERMAN  
G.E. EHRLICH (1995) LTD.  
11 MENACHEM BEGIN STREET  
RAMAT GAN, 52521  
ISRAEL

## PCT

### WRITTEN OPINION OF THE INTERNATIONAL SEARCHING AUTHORITY

(PCT Rule 43bis.1)

|   |  |  |  |
|---|--|--|--|
| Applicant's or agent's file reference<br>33103  |  | Date of mailing<br>(day/month/year) <b>25 SEP 2008</b>                     |  |
| International application No.<br>PCT/IL07/00123   |  | International filing date (day/month/year)<br>31 January 2007 (31.01.2007) |  |
| International Patent Classification (IPC) or both national classification and IPC<br>IPC: G06F 17/28( 2006.01)<br>USPC: 704/2,4,7,8 |  | Priority date (day/month/year)<br>31 January 2006 (31.01.2006)             |  |
| Applicant<br>RAMOT AT TEL AVIV UNIVERSITY LTD.  |  |  |  |

1. This opinion contains indications relating to the following items:

- ☒ Box No. I      Basis of the opinion
- ☐ Box No. II      Priority
- ☐ Box No. III      Non-establishment of opinion with regard to novelty, inventive step and industrial applicability
- ☐ Box No. IV      Lack of unity of invention
- ☒ Box No. V      Reasoned statement under Rule 43bis.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement
- ☐ Box No. VI      Certain documents cited
- ☐ Box No. VII      Certain defects in the international application
- ☐ Box No. VIII      Certain observations on the international application

#### 2. FURTHER ACTION

If a demand for international preliminary examination is made, this opinion will be considered to be a written opinion of the International Preliminary Examining Authority ("IPEA") except that this does not apply where the applicant chooses an Authority other than this one to be the IPEA and the chosen IPEA has notified the International Bureau under Rule 66.1bis(b) that written opinions of this International Searching Authority will not be so considered.

If this opinion is, as provided above, considered to be a written opinion of the IPEA, the applicant is invited to submit to the IPEA a written reply together, where appropriate, with amendments, before the expiration of 3 months from the date of mailing of Form PCT/ISA/220 or before the expiration of 22 months from the priority date, whichever expires later.

For further options, see Form PCT/ISA/220.

3. For further details, see notes to Form PCT/ISA/220.

|  |  |   |
|--|--|---|
| Name and mailing address of the ISA/ US<br>Mail Stop PCT, Attn: ISA/US<br>Commissioner for Patents<br>P.O. Box 1450<br>Alexandria, Virginia 22313-1450<br>Facsimile No. (571) 273-3201 | Date of completion of this opinion<br>02 September 2008 (02.09.2008) | Authorized officer <i>J. Roberts</i><br>LAMONT M. SPOONER<br>Telephone No. 703/308-4357 |
|--|--|---|

Form PCT/ISA/237 (cover sheet) (April 2007)

**WRITTEN OPINION OF THE  
INTERNATIONAL SEARCHING AUTHORITY**

International application No.

PCT/IL07/00123

**Box No. I Basis of this opinion**

1. With regard to the language, this opinion has been established on the basis of:
  - ☒ the international application in the language in which it was filed
  - ☐ a translation of the international application into \_\_\_\_\_, which is the language of a translation furnished for the purposes of international search (Rules 12.3(a) and 23.1(b)).
2. ☐ This opinion has been established taking into account the rectification of an obvious mistake authorized by or notified to this Authority under Rule 91 (Rule 43bis.1(a))
3. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this opinion has been established on the basis of:
  - a. type of material
    - ☐ a sequence listing
    - ☐ table(s) related to the sequence listing
  - b. format of material
    - ☐ on paper
    - ☐ in electronic form
  - c. time of filing/furnishing
    - ☐ contained in the international application as filed.
    - ☐ filed together with the international application in electronic form.
    - ☐ furnished subsequently to this Authority for the purposes of search.
4. ☐ In addition, in the case that more than one version or copy of a sequence listing and/or table(s) relating thereto has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.
5. Additional comments:

**WRITTEN OPINION OF THE  
INTERNATIONAL SEARCHING AUTHORITY**

International application No.  
PCT/IL07/00123

**Box No. V Reasoned statement under Rule 43 *bis*.1(a)(i) with regard to novelty, inventive step or industrial applicability; citations and explanations supporting such statement**

**1. Statement**

|                               |                    |     |
|-------------------------------|--------------------|-----|
| Novelty (N)                   | Claims <u>1-50</u> | YES |
|                               | Claims <u>NONE</u> | NO  |
| Inventive step (IS)           | Claims <u>NONE</u> | YES |
|                               | Claims <u>1-50</u> | NO  |
| Industrial applicability (IA) | Claims <u>1-50</u> | YES |
|                               | Claims <u>NONE</u> | NO  |

**2. Citations and explanations:**

Please See Continuation Sheet

WRITTEN OPINION OF THE  
INTERNATIONAL SEARCHING AUTHORITY

International application No.  
PCT/IL07/00123

Supplemental Box

In case the space in any of the preceding boxes is not sufficient.

**V. 2. Citations and Explanations:**

1. Claims 1-35 lack an inventive step under PCT Article 33(3) as being obvious over Nakayama et al. (Nakayama, US 5,687,383) in view of Brown et al. (Brown, US 5,293,584).

As per claims 1 and 17, Nakayama teaches a method of constructing translation rules using a first dataset of sentences in a source language and a second dataset of sentences in a target language, the translation rules being for translating the source language into the target language, the method comprising:

acquiring at least a partial source grammar characterizing the source language (Fig. 8A-his parse of source, and grammar) and at least a partial target grammar characterizing the target language (Fig. 8b-his parse of target and grammar);

for each sentence of the first dataset, generating a mapping function for mapping an ordered set of source grammar symbols being associated with said sentence to an ordered set of target grammar symbols, thereby providing a set of mapping functions (Fig. 5 item ST14-his corresponding candidates); and

archiving said set of mapping functions in a database, thereby constructing the translation rules (C.4 lines 9,10).

Nakayama does not explicitly teach translating utterances of a source language into utterances of a target language. However, Brown teaches translating utterances... (C.2 lines 11-24)

Therefore, at the time of the invention, it would have been obvious to one ordinarily skilled in the art to modify Nakayama with Brown, providing the benefit of speech input for translation.

As per claims 2 and 18, Nakayama and Brown make obvious the method of claim 1, wherein said generating said mapping function is by a machine readable dictionary (C.4 lines 5-10-his machine readable dictionary).

As per claims 3 and 19, Nakayama and Brown make obvious the method of claim 1, further comprising archiving said at least partial source grammar and said at least partial source target grammar in said database (Fig. 1, item 3 his rule/grammar dictionary).

As per claims 4 and 20, Nakayama and Brown make obvious the method of claim 1, wherein said generating said mapping function comprises, for each dataset of the first and second datasets, parsing each sentence of said dataset to provide an ordered set of symbols covering said sentence (Fig. 8A, 8B-his parse), and updating said mapping functions based on said ordered set of symbols (C.4 lines 9 and 10).

As per claims 5 and 21, Nakayama and Brown make obvious the method of claim 1, further comprising, constructing a source

**WRITTEN OPINION OF THE  
INTERNATIONAL SEARCHING AUTHORITY**

International application No.  
PCT/IL07/00123

**Supplemental Box**

In case the space in any of the preceding boxes is not sufficient.

probability matrix from the source grammar (Fig. 8A-his probability parse, Fig. 16-his matrix and score) and a target probability matrix from the target grammar, each probability matrix (Fig. 8B-his probability parse, Fig. 16-his matrix and score) comprising a plurality of entries representing co-occurrence probabilities of symbols in a respective grammar (Fig. 8A, 8B, wherein the co-occurrence of his symbol with respect to grammar is generated in the parse, Fig. 16-and generated score).

As per claims 6 and 22, Nakayama and Brown make obvious the method of claim 5, further comprising, for each dataset of the first and second datasets, parsing each sentence of said dataset to provide an ordered set of symbols covering said sentence (Fig. 8A, 8B-his parse), and updating a respective probability matrix based using said ordered set of symbols (Fig 16-updated and scored by parse and probability).

As per claims 7 and 23, Nakayama and Brown make obvious the method of claim 6, further comprising updating said set of mapping functions using said probability matrices (C.7 lines 5-14-his mapping updated based on estimation).

As per claims 8 and 24, Nakayama and Brown make obvious the method of claim 1, wherein each of said source grammar and said target grammar independently comprises:

terminals being associated with tokens of a lexicon characterizing said dataset and nonterminals being associated with equivalence classes of tokens of said lexicon and/or significant patterns of a respective dataset (Fig. 2, his terminals and nonterminals in each respective dataset, Fig. 1-his bilingual texts, C.4 lines 55-67-his original sentence and translated sentence, analyzed with respect to respective dictionary/lexicon).

As per claims 9 and 25, Nakayama and Brown make obvious the method of claim 8, wherein said acquiring said source grammar and said target grammar, comprises for each sentence of the first dataset and for each sentence of the second dataset, searching for partial overlaps between said sentence and other sentences of the respective dataset (Fig. 17-his case frame and regional determination), applying a significance test on said partial overlaps (Fig. 18-his estimation of case filler), and defining a most significant partial overlap as a significant pattern of said sentence, thereby extracting significant patterns from the first and the second datasets, thereby acquiring nonterminals for said source grammar and said target grammar (Fig 19-his extracted significant patterns).

As per claims 10 and 26, Nakayama and Brown make obvious the method of claim 8, wherein said acquiring said source grammar and said target grammar, comprises:

for each dataset of the first dataset and the second dataset: searching over said dataset for similarity sets (Fig 9 item ST21-his correspondence), each similarity set comprising a plurality of segments of size L (Fig. 8A, 8B-his J1, J2, J3 components) having L-S common tokens and S uncommon tokens (his symbol uncommon tokens), each of said plurality of segments being a portion of a different sentence of said dataset; and defining a plurality of equivalence classes corresponding to uncommon tokens of at least one similarity set (fig. 13 his "question" and "problem" similarity set from different sentences), thereby acquiring nonterminals for said source grammar and said target grammar (Fig. 14-his acquired nonterminal for question).

As per claims 11, 27, 28 and 29, Nakayama and Brown make obvious the method of claim 10, wherein said definition of said plurality of equivalence classes comprises, for each segment of each similarity set:

extracting a significant pattern corresponding to a most significant partial overlap between said segment and other segments or combination of segments of said similarity set, thereby providing, for each similarity set, a plurality of significant patterns (Fig. 17-his case information region providing overlap, and Fig. 18, his missing case for the similarity set produced in the corresponding case frame); and

using said plurality of significant patterns for classifying tokens of said similarity set into at least one equivalence class (Fig. 17-his part of speech class, see also Fig 19);

thereby defining said plurality of equivalence classes (ibid, the equivalence class identified by corresponding part of speech).

As per claims 12, 29 and 30, Nakayama and Brown make obvious method of claim 11, wherein said classification of said tokens comprises, selecting a leading significant pattern of said similarity set (Fig. 18-his estimation of missing case and case filler determined by similarity set), and defining uncommon tokens of segments corresponding to said leading significant pattern as an equivalence class (Fig. 18-his missing case information, the estimation of the case filler directly from the equivalence class).

As per claims 13 and 31, Nakayama and Brown make obvious the method of claim 11, further comprising, prior to said search for said similarity sets:

extracting a plurality of significant patterns from the dataset, each significant pattern of said plurality of significant patterns corresponding to a most significant partial overlap between one sentence of the dataset and other sentences of the dataset (Fig. 2 his file read and extracted patterns, Fig 3 his template for translation rules including overlaps, see Fig. 6, overlapping cases); and

for each significant pattern of said plurality of significant patterns, grouping at least a few tokens of said significant pattern, thereby redefining the dataset (Fig. 29-his final grouped patterns to be learned, redefining the dataset).

As per claims 14, 32 and 33, Nakayama and Brown make obvious the method of claim 13, further comprising constructing a graph having a plurality of paths representing the dataset (Fig. 8A, 8B), wherein each extraction of significant pattern is by searching for partial overlaps between paths of said graph (Fig. 9-item ST21-his case frame correspondence overlap).

**WRITTEN OPINION OF THE  
INTERNATIONAL SEARCHING AUTHORITY**

International application No.  
PCT/IL07/00123

**Supplemental Box**

In case the space in any of the preceding boxes is not sufficient.

As per claims 15 and 34, Nakayama and Brown make obvious the method of claim 14, wherein said graph comprises a plurality of vertices (Fig. 8A, 8B, his J1, J2, J3, E1, E2, E3), each representing one token of the lexicon (ibid), and further wherein each path of said plurality of paths comprises a sequence of vertices respectively corresponding to one sentence of the dataset (Fig 8A, 8B-his sequence of sentence from Figs. 7A, 7B).

As per claims 16 and 35, Nakayama and Brown make obvious the method of claim 14, further comprising calculating, for each path, a set of probability functions characterizing said partial overlaps (Fig. 18 item ST33-his probability/estimation of case filler, based on paths/overlap).

2. Claims 36-50 lack an inventive step under PCT Article 33(3) as being obvious over Nakayama et al. (Nakayama, US 5,687,383) in view of Brown et al. (Brown, US 5,293,584).

Claims 36-50 lack inventive step under PCT Article 33(3) as being obvious over Brown et al. (Brown, US 5,293,584).

As per claims 36 and 42, Brown teaches a method of translating an utterance from a source language to a target language, comprising:

employing a structured stochastic language model so as to generate from the utterance a plurality of candidate utterances in the target language (Fig. 2 item 20-his candidate word generator), and so as to assign a score to each candidate utterance of said plurality of candidate utterances (Fig. 1 item 14, item 24); and

selecting a candidate utterance having an optimal score, thereby translating the utterance from the source language to the target language (C.2 lines 32-36, 41-47-his best hypothesis match score translation).

As per claims 37 and 43, Brown teaches the method of claim 36, further comprising accessing a database to obtain at least a partial source grammar characterizing the source language (C.2 lines 1-10-his source language model) and at least a partial target grammar characterizing the target language (C.2 lines 1-10-his target/translation model), wherein said generation of said plurality of candidate utterances is based on said at least partial grammars (ibid, C.2 lines 48-65-his candidate utterances based on the models).

As per claims 38 and 44, Brown teaches the method of claim 37, further comprising, prior to said generation of said plurality of candidate utterances:

accessing said database to obtain translation rules defined by a set of mapping functions for mapping ordered sets of source grammar symbols to ordered sets of target grammar symbols (C.2 lines 5-7-his source language information, mapped to speech hypothesis);

parsing the utterance to obtain an ordered set of source grammar symbols covering the utterance (C.2 lines 14-18-his sequence of coded utterance), thereby providing an utterance cover in said at least partial source grammar (ibid);

using said translation rules for mapping said utterance cover to an ordered set of target grammar symbols (C.2 lines 32-33-his translation match mapping for the sequence above); and

assigning prior probabilities in said structured stochastic language model for each target grammar symbol of said grammar symbols set (C.2 lines 48-54-his candidate target probabilities).

As per claims 39 and 45, Brown teaches the method of claim 38, further comprising prior to said selection of said candidate utterance, processing said plurality of candidate utterances according to additional ranking criteria (C.2 lines 41-46-his best estimate of candidate utterance).

As per claims 40 and 46, Brown teaches the method of claim 39, wherein said processing comprises, for each candidate utterance of said plurality of candidate utterances:

parsing said candidate utterance to obtain an ordered set of target grammar symbols covering said candidate utterance C.2 lines 14-18-his sequence of coded utterance), thereby providing a candidate utterance cover in said at least partial target grammar (C.2 lines 41-46-his best estimate of candidate utterance as the cover); and

performing correspondence analysis to provide goodness of correspondence between said candidate utterance cover in said at least partial target grammar and said utterance cover in said at least partial source grammar (C.2 lines 55-65-his translation match score for the speech hypothesis as the correspondence analysis).

As per claims 41 and 47, Brown teaches the method of claim 38, wherein said employing said stochastic language model comprises replacing said prior probabilities with conditional probabilities representing a discourse context of the utterance (C.6 lines 20-25-his conditional translation match, and P(T'), as replacing P(T), (T') as generated from the language match score generator).

As per claim 48, Brown teaches a text processing system having a translator, the translator comprising the apparatus of any of claims 42-47 (C.2 lines 5-10-his translation).

As per claim 49, Brown teaches a text processing system having a style checker, the style checker comprising the apparatus of any of claims 42-47 (C1 lines 64-67, his language complexity checker).